

---

# Local and Dimension Adaptive Sparse Grid Interpolation and Quadrature

J.D Jakeman and S.G. Roberts

<sup>1</sup> J.D. Jakeman, Department of Mathematics, Purdue University, West Lafayette, IN USA  
47907, jjakeman@purdue.edu

<sup>2</sup> S.G. Roberts, Mathematical Sciences Institute, Australian National University, ACT  
Australia 0200, stephen.roberts@anu.edu.au

**Summary.** In this paper we present a locally and dimension-adaptive sparse grid method for interpolation and integration of high-dimensional functions with discontinuities. The proposed algorithm combines the strengths of the generalised sparse grid algorithm and hierarchical surplus-guided local adaptivity. A high-degree basis is used to obtain a high-order method which, given sufficient smoothness, performs significantly better than the piecewise-linear basis. The underlying generalised sparse grid algorithm greedily selects the dimensions and variable interactions that contribute most to the variability of a function. The hierarchical surplus of points within the sparse grid is used as an error criterion for local refinement with the aim of concentrating computational effort within rapidly varying or discontinuous regions. This approach limits the number of points that are invested in ‘unimportant’ dimensions and regions within the high-dimensional domain. We show the utility of the proposed method for non-smooth functions with hundreds of variables.

## 1 Introduction

The need for interpolation and integration of high-dimensional functions arises in many fields including finance, physics, chemistry, and uncertainty quantification. Sparse grids have emerged as an extremely useful tool to construct such multi-dimensional approximations. They have been extensively used for high-dimensional interpolation [1, 5] and quadrature [2, 6] and have been shown, under certain conditions, to obtain significantly higher rates of convergence than many existing methods. For example, the complexity of the Monte Carlo method with  $n$  samples is  $O(n^{-1/2})$ , whereas the complexity of the sparse grid method [13] is  $O(n^{-r}(\log n)^{(d-1)(r+1)})$  when used to approximate  $d$ -dimensional integrands which have bounded mixed partial derivatives of order  $r$ . Sparse grids achieve faster rates of convergence by taking advantage of higher smoothness and lower-effective dimensionality of the integrand.

Standard sparse grids are isotropic, treating all dimensions equally. Although an advance on alternative methods, such approximations can still be improved. Many problems vary rapidly in only some dimensions, remaining less variable in other dimensions. Consequently it is advantageous to increase the level of accuracy only

in certain highly varying dimensions, resulting in so-called adaptive or anisotropic grids. In some cases the important dimensions can be determined a priori, but in most cases the grid points must be chosen during the computational procedure.

Gerstner and Griebel [6] developed a dimension-adaptive tensor-product quadrature method to approximate high-dimensional functions by a sum of lower-dimensional terms. The method is based upon a generalisation of the traditional isotropic sparse grid index set that, given appropriate error estimators, can automatically concentrate computational effort in important dimensions. Recently Griebel and Hotlz [8] developed a new general class of dimension-adaptive quadrature methods. The quadrature schemes detect and exploit the low effect dimensionality of a function. This is achieved by truncation and discretization of the anchored-ANOVA decomposition of the function being approximated. This method has been used successfully to estimate high-dimensional integrals arising in finance.

In addition to dimension based adaptivity, efficiency in approximation of high-dimensional functions can also be obtained through local adaptation. Locally adaptive sparse grids were first used by Griebel [7] to solve the solution of partial differential equation and have also been used to quantify uncertainty in mathematical models [9], interpolate and integrate functions [12] and scattered data approximation problem [11]. The refinement of the sparse grid is guided by the magnitude of the so-called hierarchical surplus, which is the difference between the true function and the approximation at a new grid point before that point is used in the interpolation. When a grid point is identified for refinement, new points are inserted locally around that point in every dimension. These new points are subsequently refined. The resulting method automatically concentrates function evaluations in rapidly varying or discontinuous regions.

The aforementioned locally adaptive sparse grid methods are implicitly dimension-adaptive, but often points are constructed unnecessarily in ‘unimportant’ dimensions. In comparison, the generalised sparse grid algorithm [6] provides an efficient means of identifying the effective dimensionality of a problem and restricting function evaluations to that sub-dimensional space. This method performs extremely well when the solution is smooth. However the efficiency of the generalised sparse grid method can be significantly improved when only small regions of the input space contribute to the model’s variability.

In this paper we combine local refinement with the dimension-adaptive algorithm of the generalised sparse grid method to construct an efficient high-dimensional interpolation method. Furthermore we utilise the localised polynomial basis proposed by Bungartz [3] to create a higher-order method which achieves fast rates of convergence in smooth regions and accuracy, comparable to linear methods, around discontinuities. We coin this approach the *h*-Adaptive Generalised Sparse Grid (*h*-GSG) method. The convergence of the proposed method is analysed, with respect to the order of the local polynomial basis and the dimensionality of the input space.

## 2 Adaptive sparse grids

In this paper we will attempt to interpolate and integrate functions  $f : \Omega \rightarrow \mathbb{R}$  defined on a  $d$ -dimensional bounded domain  $\Omega$ . We need not know the closed form of  $f$ , we only require that the function  $f$  can be evaluated at arbitrary points in  $\Omega$  using a numerical code.

### 2.1 Interpolation

To construct an interpolant of  $f$ , we must first discretize  $\Omega$ . Without loss of generality let us consider functions defined on the  $d$ -dimensional unit hypercube  $\Omega = [0, 1]^d$ . Sparse grids are a direct sum of anisotropic grids  $\Omega_{\mathbf{i}}$  on the domain  $\Omega$  where  $\mathbf{i} = (i_1, \dots, i_d) \in \mathbb{N}^d$  is a multi-index denoting the level of refinement of the grid in each dimension  $d$ . Each grid  $\Omega_{\mathbf{i}}$  is a tensor product of one-dimensional grids

$$\Omega_{\mathbf{i}} = (x_{i,1}, \dots, x_{i,m_i}) \quad (1)$$

where  $m_i$  is odd and represents the number of points  $x_{i,j}$  in the  $i$ th level one dimensional grid. Specifically

$$\Omega_{\mathbf{i}} = \bigotimes_{n=1}^d \Omega_{i_n}$$

which consists of the points  $\mathbf{x}_{\mathbf{i},\mathbf{j}} = (x_{i_1,j_1}, \dots, x_{i_d,j_d})$  and where  $\mathbf{i}$  indicates the level of refinement and  $\mathbf{j}$  denotes the location of a given grid point. The exact coordinates of each point and the total number of points  $m_{i_1} \times \dots \times m_{i_d}$  is dependent on the type of one-dimensional grids used.

Each grid  $\Omega_{\mathbf{i}}$  is associated with a discrete approximation space  $V_{\mathbf{i}}$  and a set of basis functions that span the discrete space. The types of basis functions that can be used are dependent on the type of one-dimensional grids employed. The most frequently used and simplest choice are the multi-linear piecewise basis functions [4, 7, 9], based upon the one-dimensional formula

$$\Psi_{i,j}(x) = \begin{cases} 1 - (m_i - 1)|x - x_{i,j}| & \text{if } |x - x_{i,j}| < h_i \\ 0 & \text{otherwise} \end{cases}$$

centered at the points

$$x_{i,j} = \begin{cases} j \times h & i \geq 1 \text{ and } 0 \leq j \leq m_i \\ 0.5 & i = 1 \end{cases}$$

where  $h_i = 1/(m_i - 1)$ .

These 1D basis functions can be used to form a set of  $d$ -dimensional basis functions

$$\Psi_{\mathbf{i},\mathbf{j}}(\mathbf{x}) = \prod_{n=1}^d \Psi_{i_n,j_n}(x_n)$$

which span the discrete space  $V_{\mathbf{i}}$ . Specifically

$$V_{\mathbf{i}} = \text{span} \{ \Psi_{\mathbf{i}, \mathbf{j}} \mid j_n = 1, \dots, m_{i_n}, n = 1, \dots, d \} \quad (2)$$

The spaces  $V_{\mathbf{i}}$  can be used to define hierarchical difference spaces  $W_{\mathbf{i}}$

$$W_{\mathbf{i}} = V_{\mathbf{i}} \setminus \bigoplus_{n=1}^d V_{\mathbf{i}-\mathbf{e}_n} \quad (3)$$

These spaces consist of all the basis functions  $\Psi_{\mathbf{i}, \mathbf{j}} \in V_{\mathbf{i}}$  with associated points  $\mathbf{x}_{\mathbf{i}, \mathbf{j}}$  that are not associated with any of the basis functions in spaces smaller than  $V_{\mathbf{i}}$ . A discrete space  $V_{\mathbf{k}}$  is smaller than a space  $V_{\mathbf{i}}$  if  $\mathbf{k} \leq \mathbf{i}$ . Setting  $V_{\mathbf{i}} = 0$  and using (2) and (3) we obtain

$$W_{\mathbf{i}} = \text{span} \{ \Psi_{\mathbf{i}, \mathbf{j}} \mid \mathbf{j} \in B_{\mathbf{i}} \}$$

where

$$B_{\mathbf{i}} = \{ \mathbf{j} : j_n = 1, \dots, m_{i_n}, j \text{ odd } n = 1, \dots, d \} \quad (4)$$

These hierarchical difference spaces can be used to decompose the input space  $\Omega = V$  such that

$$V = \bigoplus_{k_1=0}^{\infty} \dots \bigoplus_{k_d=0}^{\infty} W_{\mathbf{k}} = \bigoplus_{\mathbf{k} \in \mathbb{R}^d} W_{\mathbf{k}}$$

For numerical purposes we must truncate the number of difference spaces used to construct  $V$  to some level  $l$ . Specifically the classical finite dimensional sparse grid space is defined by

$$V_{l,d}^{(1)} = \bigoplus_{|\mathbf{i}|_1 \leq l} W_{\mathbf{i}} \quad (5)$$

With such a decomposition any function  $f(\mathbf{x}) \in V$  can be approximated by

$$f_{l,d}(\mathbf{x}) = \sum_{|\mathbf{i}|_1 \leq l} \sum_{\mathbf{j} \in B_{\mathbf{i}}} v_{\mathbf{i}, \mathbf{j}} \Psi_{\mathbf{i}, \mathbf{j}}(\mathbf{x}) \quad (6)$$

where  $v_{\mathbf{i}, \mathbf{j}} \in \mathbb{R}$  are the coefficient values of the hierarchical product basis, also known as the hierarchical surplus.

The size of the sparse grid space is  $|V_{l,d}^{(1)}| = \mathcal{O}(h_l^{-l} \cdot |\log_2 h_l|^{d-1}) = \mathcal{O}(2^l \cdot l^{d-1})$  which is a significant reduction on the  $\mathcal{O}(2^{l \cdot d})$  number of points required by the full tensor product space  $V_{l,d}^{(\infty)}$  obtained by choosing  $|\mathbf{i}|_{\infty} = \max_{0 \leq n \leq d} i_n \leq l$ .

## 2.2 Quadrature

The extension from interpolation to quadrature is straightforward. We can approximate the integral of a function  $f$

$$I[f(\mathbf{x})] = \int_{I_{\mathbf{x}}} f(\mathbf{x}) d\mu(\mathbf{x})$$

using the hierarchical sparse grid interpolant (6). Utilizing this formulation these integrals can be approximated by

$$\begin{aligned} I[f_{l,d}(\mathbf{x})] &= \int_{I_x} \sum_{|\mathbf{i}|_1 \leq l} \sum_{\mathbf{j} \in B_{\mathbf{i}}} v_{\mathbf{i},\mathbf{j}} \Psi_{\mathbf{i},\mathbf{j}}(\mathbf{x}) d\mu(\mathbf{x}) \\ &= \sum_{|\mathbf{i}|_1 \leq l} \sum_{\mathbf{j} \in B_{\mathbf{i}}} v_{\mathbf{i},\mathbf{j}} w_{\mathbf{i},\mathbf{j}} \end{aligned}$$

where the weights

$$w_{\mathbf{i},\mathbf{j}} = \int_{I_x} \Psi_{\mathbf{i},\mathbf{j}}(\mathbf{x}) d\mu(\mathbf{x})$$

can be calculated easily and with no need for extra function evaluations once the interpolant has been constructed. One simply needs to store the volumes of the high-order basis functions. For  $d\mu = d\mathbf{x}$  these volumes can be calculated analytically.

### 2.3 A Local High-Order Basis

Sparse grids are not restricted to piecewise multi-linear basis functions that are constructed on equidistant grids. Various formulations exist. In the following we propose a high order local basis for interpolation and quadrature. This basis was first proposed by Bungartz [3] for the solution of partial differential equations. The local nature of the basis functions allows for local adaptivity and restricts the effects of Gibbs type phenomenon experienced by global polynomial approximation whilst still achieving polynomial convergence in smooth regions.

As with the linear case, we restrict our attention to grids  $\Omega_i$  with mesh spacing that is equidistant with respect to each individual dimension but may vary between dimensions. Let  $\Psi_{i,j}^{(p)}$  denote a one-dimensional polynomial of degree  $p$  defined on the interval  $[x_{i,j} - h_i, x_{i,j} + h_i]$ . This localized support is essential for application of the high-order basis to non-smooth problems and the ultimate goal of an adaptive method. Uniquely defining this basis requires  $p+1$  conditions that  $\Psi_{i,j}^{(p)}$  must satisfy.

Here we take advantage of the fact that each point  $\mathbf{x}_{i,j}$  has an ancestry. The one-dimensional equidistant points  $x_{i,j}$  can be considered as a tree-like data structure. The coordinate of each point is defined uniquely by the level  $i$  and the position  $j$ . With this observation we can define a local  $p$ -th order polynomial using the points  $x_{i,j} - h_i, x_{i,j}, x_{i,j} + h_i$  and the next  $p-2$  closest hierarchical ancestors of  $x_{i,j}$ .

**Definition 1.** *Given the one-dimensional grid  $\Omega_i$  with grid points defined according to (2.1) the  $p$ -th order basis function  $\Psi_{i,j}^{(p)}$  is the hierarchical interpolant of the point  $x_{i,j} - h_i, x_{i,j}, x_{i,j} + h_i$  and the next  $p-2$  closest hierarchical ancestors of  $x_{i,j}$  restricted to the local support  $[x_{i,j} - h_i, x_{i,j} + h_i]$ . Specifically by renaming all the points except  $x_{i,j}$  in ascending order as  $x_0, \dots, x_p$  the piecewise Lagrange basis can be written*

$$\Psi_{i,j}^{(p)}(x) = \begin{cases} \prod_{k=0}^p \frac{x - x_k}{x_{i,j} - x_k} & \text{if } |x - x_{i,j}| < h_i \\ 0 & \text{otherwise} \end{cases}$$

The order  $p$  of the basis function is dependent on the hierarchical level  $i$  of  $x_{i,j}$ . For  $p > 2$ ,  $p+1$  ancestors are needed to construct the basis  $\Psi_{i,j}$ . On level one, only one ancestor ( $x = 0.5$ , on level zero) is available and thus only linear basis functions can be used. On level two, only two ancestors exist, and thus only linear or quadratic basis functions can be used. Subsequently basis functions of degree  $p$  can only be used when  $i \geq p$ . This represents a slight modification of the approach employed by Bungartz [3] who allowed basis functions of degree  $p$  to be used when  $i \geq p-1$ . Bungartz approach was designed for sparse grids with homogeneous boundary conditions.

Throughout the remainder of this manuscript we restrict our attention to basis functions with fixed maximum degree  $p_{\max}$ . That is, the order of the basis is increased with each level of the sparse grid until the order of the basis is  $p_{\max} - 1$ . On all subsequent levels the order of the basis is kept constant at  $p_{\max}$ . The tensor product construction of the multi-dimensional basis means that the degree  $\mathbf{p} = (p_1, \dots, p_d)$  of a  $d$ -dimensional basis function  $\Psi_{\mathbf{i},\mathbf{j}}$  must satisfy

$$0 \leq p_n = \min\{p_{\max}, i_n\}, \quad i_n \geq 0 \quad , n = 1, \dots, d$$

Here  $p_n = 0$  represents the constant function centred at the midpoint of  $[0, 1]$ .

## 2.4 Adaptivity

The classical sparse grids presented in Section 2.1 are based upon the index set

$$\mathcal{I} = \{\mathbf{i} \in \mathbb{N}^d : |\mathbf{i}|_1 \leq l\}$$

This construction delays the curse of dimensionality by assuming that the importance of any interaction between a subset of a function's variables decreases as the number of variables involved in the interaction (interaction order) increases. Although an advance on full tensor product spaces, such approximations can still be improved. The classical sparse grid construction treats all dimensions equally and all interactions of the same order equally. In practice, often only a small subset of variables and interactions contributes significantly to the variability of the function  $f$ . Moreover, frequently only small regions within the input space possess high variability. In some cases the important dimensions, interactions and regions can be determined a priori, but in most cases these properties must be identified during the computational procedure.

The generalised sparse grid method [6] is extremely effective at determining the dimensions and interactions that contribute significantly to the function variability, according to some predefined measure. However the efficiency of this method deteriorates when a large proportion of the function variability is concentrated in small regions of the input space. In contrast to the generalised sparse grid method, locally-adaptive methods, such as that of Ma and Zabaras [9], attempt to reduce the number of points in a sparse grid by concentrating refinement only in rapidly varying or discontinuous regions. This method is implicitly dimension-adaptive but often points are constructed necessarily in 'unimportant' dimensions [10].

In this section we propose a method which combines the strengths of both local-adaptivity and the generalised sparse grid algorithm. We coin this approach the  $h$ -Adaptive Generalised Sparse Grid ( $h$ -GSG) method.

### Generalised Sparse Grid (GSG) Algorithm

Gerstner [6] generalised the sparse grid construction by considering the index sets based upon the admissibility criterion

$$\mathbf{i} - \mathbf{e}_j \in \mathcal{I} \text{ for } 1 \leq j \leq d, i_j > 1 \quad (7)$$

This so-called generalised sparse grid method [6] is extremely effective at determining the hierarchical difference spaces that contribute significantly to the function variability, according to some predefined measure.

The generalised sparse grid method is a greedy algorithm which attempts to find the index set  $\mathcal{I}$  such that for a given number of points the approximation error is minimized. Starting with  $\mathcal{I} = \{\mathbf{0}\}$  the index set is built iteratively by searching the forward neighbourhood of the current index set for new admissible indices. The forward neighbourhood of an index  $\mathbf{i}$  is the set of  $d$  indices  $\{\mathbf{i} + \mathbf{e}_j : 1 \leq j \leq d\}$ . Similarly, the backwards neighbourhood is just  $\{\mathbf{i} - \mathbf{e}_j : 1 \leq j \leq d\}$ .

Once the forward neighbourhood has been identified, each forward neighbour is checked for admissibility using (7). The grid points associated with each admissible index are then evaluated and the error of these spaces calculated. The calculation of these errors will be addressed shortly. The forward neighbour with the largest error is then added to the current index set  $\mathcal{I}$  and the set of admissible indices is updated.

To facilitate easy computation of new admissible indices we partition the index set  $\mathcal{I}$  into two disjoint sets  $\mathcal{O}$  and  $\mathcal{A}$ , which Gerstner [6] refers to as the old and active index sets, respectively. The active index set  $\mathcal{A}$  contains all the indices in  $\mathcal{I}$  that have been constructed but whose forward neighbours have not been considered. The old index set contains all the indices remaining in the current index set  $\mathcal{I}$ . The algorithm proceeds by searching the forward neighbourhood of the index  $\mathbf{i} \in \mathcal{A}$  with the largest error for admissible indices. All the new admissible index sets are added to the active index set  $\mathcal{A}$  and the index  $\mathbf{i}$  is then added to the old index set  $\mathcal{O}$ . This process is repeated until a global error is below a predefined tolerance  $\varepsilon$ .

The exact error associated with each index  $\mathbf{i}$  is unknown. Consequently each time an index  $\mathbf{i}$  is deemed admissible an approximation of the error  $r_{\mathbf{i}}$  must be used. Numerous error criteria can be utilised. Here we employ the error measure

$$r_{\mathbf{i}} = \left| \sum_{\mathbf{j} \in B_{\mathbf{i}}} v_{\mathbf{i}, \mathbf{j}} \cdot w_{\mathbf{i}, \mathbf{j}} \right|$$

These index-based error criteria can be used to approximate the global error. We propose the following global error indicator  $r$

$$r = \sum_{\mathbf{i} \in \mathcal{A}} r_{\mathbf{i}}$$

When  $r < \varepsilon$  the generalised sparse grid algorithm is terminated. The generalised sparse grid algorithm is outlined in Algorithm 1.

---

**Algorithm 1** Generalised Sparse Grid Approximation

---

```

i = (0, ..., 0)
A := {i}
O :=  $\emptyset$ 
r :=  $r_i$ 
while  $r > \varepsilon$  do
    select i  $\in \mathcal{A}$  with largest error indicator  $r_i$ 
    A :=  $\mathcal{A} \setminus \{\mathbf{i}\}$ 
    O :=  $\mathcal{O} \cup \{\mathbf{i}\}$ 
    r :=  $r - r_i$ 
    for  $k := 1, \dots, d$  do
        j = i + e $k$ 
        if j - e $n$   $\in \mathcal{O}$   $\forall n = 1, \dots, d$  then
            A :=  $\mathcal{A} \cup \{\mathbf{j}\}$ 
            CreateGrid(j)
            r :=  $r + r_j$ 
        end if
    end for
end while

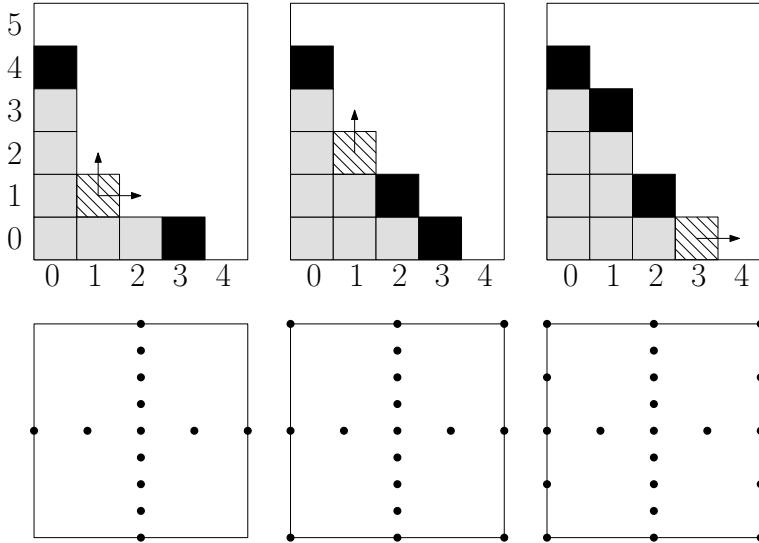
```

---

Three steps of the generalised sparse grid algorithm depicting the construction of the sparse grid index set are shown in Figure 1. The top row represents the current index sets. The bottom row depicts the corresponding sparse grid. At each step the forward neighbours of the grid index **i** with the largest error  $r_i$  (striped box) are checked for admissibility. A forward neighbour is admissible if all indices in its backwards neighbourhood are in the old index set (grey boxes). All admissible indices (pointed to by an arrow) are added to the active index set (black and striped boxes).

The striped box **i** = (1, 1) in the first step has two admissible neighbours as the backwards neighbourhoods of both forward neighbours are complete. In comparison the striped box **i** = (2, 1) in the second step only has one admissible index. The index **j**<sub>1</sub> = (3, 1) has two backwards neighbours **j**<sub>1</sub> - **e**<sub>1</sub> = (0, 2) and **j**<sub>1</sub> - **e**<sub>2</sub> = (1, 1) in the old index set, and thus is admissible. In contrast the index **j**<sub>2</sub> = (2, 2) has one backwards neighbour in the old index set **j**<sub>2</sub> - **e**<sub>1</sub> = **j**<sub>2</sub> and one in the active set **j**<sub>2</sub> - **e**<sub>2</sub> = (2, 1), and so is not admissible.

Although we wish to use the generalised sparse grid algorithm for interpolation the error criteria we have proposed are based upon an integral formulation. Specifically the error indicator  $r_i$  measures the contribution of the index **i** to the global integral approximation. Furthermore the algorithm is terminated when the approximated error  $r$  in the integral is below a predefined threshold  $\varepsilon$ . This choice was made purposefully.

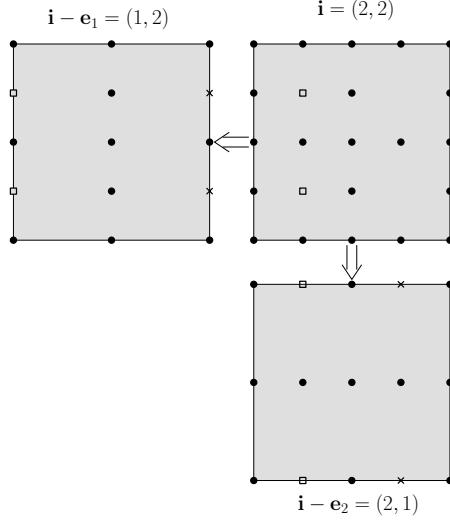


**Fig. 1.** Three steps of the generalised sparse grid algorithm. The top row represents the current index sets. Active grid indices are in black, indices in the old index set  $\mathcal{O}$  are in grey and the active index with the largest error indicator is striped. The bottom row depicts the corresponding sparse grid. Only the points associated with indices in the old index set are shown.

The magnitude of the hierarchical surplus, which is the size of the difference between the true function and the sparse grid approximation at a grid point, may be more synonymous with interpolation. Simply adding indices with large hierarchical surpluses, however, is inefficient. The magnitude of the hierarchical surpluses decays slowly in regions adjacent to discontinuities. At the site of jump discontinuities the hierarchical surplus will be at best half of the magnitude of the jump, for any finite number of grid points. Thus the algorithm can proceed much further than is necessary. The use of the error criterion  $r_i$  provides a lower bound on the size of the support of the basis functions used by weighting the magnitude of the hierarchical surplus by the probability that an arbitrary point  $x$  will fall within its support.

### Regional Adaptivity

Each time a grid index  $i$  is added to the active index set, the traditional generalised sparse grid algorithm evaluates all the points in the set  $B_i$ . Such an approach is inefficient if a large proportion of the function variability is concentrated in small regions of the input space. When using equidistant grids, the creation of the grid  $i$  requires approximately two times the number of grid points (and thus function evaluations) than those necessary to construct the index  $i - e_j$ . Consequently we propose introducing a locally adaptive procedure to construct the points associated with each grid index. To incorporate local adaptivity into the generalised sparse grid algorithm we



**Fig. 2.** An example of local adaptation integrated with the generalised sparse grid algorithm. Assume that the function only varies significantly in the left half of the domain. The top left and bottom right grids are the backwards neighbours of the grid being created. Circles represent points in the sparse grid, squares are points in the active point sets, and crosses are points in the redundant point sets. The active (square) points in the backwards neighbours are refined to produce the set of new points that must be added. In this example only two new points (squares in grid  $\mathbf{i} = (2, 2)$ ) are added.

define the two sets  $\mathcal{A}_i$  and  $\mathcal{R}_i$  for each grid index  $\mathbf{i}$ . We refer to these sets respectively as the active point set and redundant point set of the grid index  $\mathbf{i}$ . The active point set  $\mathcal{A}_i$  contains all admissible points associated with the index  $\mathbf{i}$  with an error indicator  $\gamma_{i,j} \geq \varepsilon$ . The redundant point set  $\mathcal{R}_i$  contains all admissible points with  $\gamma_{i,j} < \varepsilon$ . A point is admissible if one of its  $d$  possible ancestors exists in the grids associated with the backwards neighbourhood of  $\mathbf{i}$ . If, and only if, a grid point is admissible it is created (the function is evaluated) and the error indicator  $\gamma_{i,j}$  calculated. This drastically reduces the number of points generated when a new grid index is created.

To guide local refinement we propose using the error indicator

$$\gamma_{i,j} = |v_{i,j} \cdot w_{i,j}| \quad (8)$$

If  $\gamma_{i,j} \geq \varepsilon$  the point  $\mathbf{x}_{i,j}$  is added to the active point set  $\mathcal{A}_i$ , otherwise it is added to the redundant index set  $\mathcal{R}_i$ . The procedure used to implement  $h$ -adaptivity on each grid index is outlined in Algorithm 2.

Figure 2 shows an example of  $h$ -adaptivity integrated with the generalised sparse grid algorithm. Here the grid index  $\mathbf{i} = (2, 2)$  has been deemed admissible by the generalised sparse grid algorithm. Both the backwards neighbours ( $\mathbf{i} - \mathbf{e}_1 = (1, 2)$  and  $\mathbf{i} - \mathbf{e}_2 = (2, 1)$ ) exist in the old index set  $\mathcal{O}$ . The active points in the backwards neighbours are used to determine which points in the active index  $\mathbf{i}$  should be eval-

**Algorithm 2** CreateGrid( $\mathbf{i}$ )

---

```

for ( $n \in \{1, \dots, d\}$ ) do
  for ( $\mathbf{x}_{\mathbf{i}-\mathbf{e}_n, \mathbf{j}} \in \mathcal{A}_{\mathbf{i}-\mathbf{e}_n}$ ) do
     $\mathcal{C} = \text{FindAxialChildren}(\mathbf{x}_{\mathbf{i}-\mathbf{e}_n, \mathbf{j}}, n)$ 
    for ( $\mathbf{x}_{\mathbf{i}, \mathbf{k}} \in \mathcal{C}$ ) do
      if ( $\gamma_{\mathbf{i}, \mathbf{k}} \geq \varepsilon$ ) then
         $\mathcal{A}_{\mathbf{i}} := \mathcal{A}_{\mathbf{i}} \cup \{\mathbf{x}_{\mathbf{i}, \mathbf{k}}\}$ 
      else
         $\mathcal{R}_{\mathbf{i}} := \mathcal{R}_{\mathbf{i}} \cup \{\mathbf{x}_{\mathbf{i}, \mathbf{k}}\}$ 
      end if
    end for
  end for
end for

```

---

uated. For any point in the active set of the  $n$ -th backwards neighbour  $n = 1, \dots, d$  the children of that point are created in the  $n$ -th axial direction. This refinement is carried out for all points in the active point set of the backward neighbour and for all backwards neighbours.

### Efficient Termination

The generalised sparse grid (GSG) algorithm is a greedy algorithm which efficiently identifies the sparse grid index set that is necessary to interpolate a function up to a level of predefined accuracy. The algorithm can determine the number of variable interactions and the individual importance of each variable [8]. This is achieved by successively adding the grid index with the largest error indicator to the old index set and searching its forward neighbourhood for admissible indices. Every admissible index is added (and thus created) to the active index set without regard for the error associated with that grid. The algorithm finally terminates when  $\sum_{\mathbf{i} \in \mathcal{A}} r_{\mathbf{i}} < \varepsilon$ .

The decision to add all indices  $\mathbf{i}$ , regardless of the size of their associated error indicator  $r_{\mathbf{i}}$ , typically results in the creation of a large number of grids with  $r_{\mathbf{i}} \ll \varepsilon$  and which have little effect on the accuracy of the approximation. To reduce the number of these unimportant indices we propose only adding admissible indices with  $r_{\mathbf{i}} \geq \varepsilon$  to the active index set  $\mathcal{A}$ . This significantly reduces the number of grid indices in the final index set  $\mathcal{I}$  and thus the total number of function evaluations, with only minor effect on the overall accuracy of the generalised sparse grid method.

## 3 Error Analysis

In this section we derive a bound on the error of the proposed  $h$ -GSG method. For ease of discussion let us rewrite (6) in the following form

$$f_{l,d}(\xi) = \sum_{\mathbf{i} \leq l} f_{\mathbf{i}}(\xi), \quad f_{\mathbf{i}}(\xi) = \sum_{\mathbf{j} \in B_{\mathbf{i}}} v_{\mathbf{i}, \mathbf{j}}^{(p)} \cdot \Psi_{\mathbf{i}, \mathbf{j}}^{(p)} \in W_{\mathbf{i}} \quad (9)$$

The proposed  $h$ -GSG algorithm terminates when all points in the sparse grid with an error indicator  $\gamma_{i,j} \geq \varepsilon$  have been considered. This truncation of the sparse grid space has an effect on the accuracy of the approximation. This effect is quantified by the following theorem.

**Theorem 1.** *Let  $f_{\varepsilon,\text{opt}}$  be an interpolation of  $u$  that obtains  $\|f - f_{\varepsilon,\text{opt}}\|_q \leq \varepsilon$  with the least number of function evaluations. Then for any function  $u$  and a given tolerance  $\varepsilon > 0$  and the error criterion  $\gamma_{i,j} = \|v_{i,j} \cdot \Psi_{i,j}\|_q$ , the  $h$ -GSG approximation  $f_{\varepsilon,d}$  satisfies*

$$\|f - f_{\varepsilon,d}\|_q \leq \varepsilon (1 + N(\varepsilon))$$

where  $N(\varepsilon)$  is the number of points in the optimal interpolant but not in the  $h$ -GSG interpolant.

*Proof.* Let

$$f_{\varepsilon,\text{opt}} = \sum_{(i,j) \in P_{\varepsilon,\text{opt}}} v_{i,j} \cdot \Psi_{i,j}$$

be an interpolation of  $f$  that obtains  $\|f - f_{\varepsilon,\text{opt}}\|_q \leq \varepsilon$ . The points in this optimal approximant are defined by the index set

$$P_{\varepsilon,\text{opt}} := \{(i,j) : i \in \mathcal{I}_{\varepsilon,\text{opt}} \text{ and } j \in B_{i,\varepsilon,\text{opt}} \subseteq B_i\}$$

Similarly denote the  $h$ -GSG interpolant with an error indicator  $\gamma_{i,j}$  by

$$f_{\varepsilon,d} = \sum_{(i,j) \in P_{\varepsilon,d}} v_{i,j} \cdot \Psi_{i,j}$$

where the point indices in the  $h$ -GSG approximant are

$$P_{\varepsilon,d} := \{(i,j) : i \in \mathcal{I}_{\varepsilon,d}, j \in B_i \text{ and } |v_{i,j}| \geq \varepsilon\}$$

Now denote  $P^{\text{common}} := (P_{\varepsilon,d} \cap P_{\varepsilon,\text{opt}})$  the set of indices common to both the optimal and  $h$ -GSG approximants and denote  $P^{\text{unique}} := ((P_{\varepsilon,d} \cup P_{\varepsilon,\text{opt}}) \setminus P^{\text{common}})$  the set of indices that exist only in  $P_{\varepsilon,\text{opt}}$  or  $P_{\varepsilon,d}$ .

We can split  $P^{\text{unique}}$  further into  $P_{\varepsilon,\text{opt}}^{\text{unique}}$  and  $P_{\varepsilon,d}^{\text{unique}}$  which are points unique to the optimal approximant and the  $h$ -GSG approximants respectively. Using this splitting and the linearity of the hierarchical interpolants  $f_{\varepsilon,\text{opt}}$  and  $f_{\varepsilon,d}$ , yields

$$f_{\varepsilon,\text{opt}} = f_{\varepsilon,\text{opt}}^{\text{unique}} + f_{\varepsilon,\text{opt}}^{\text{common}} \quad \text{and} \quad f_{\varepsilon,d} = f_{\varepsilon,d}^{\text{unique}} + f_{\varepsilon,d}^{\text{common}}$$

where

$$f_{\varepsilon,\text{opt}}^{\text{unique}} = \sum_{(i,j) \in P_{\varepsilon,\text{opt}}^{\text{unique}}} v_{i,j} \cdot \Psi_{i,j} \quad , \quad f_{\varepsilon,d}^{\text{unique}} = \sum_{(i,j) \in P_{\varepsilon,d}^{\text{unique}}} v_{i,j} \cdot \Psi_{i,j}$$

and

$$f_{\varepsilon,\text{opt}}^{\text{common}} = \sum_{(i,j) \in P_{\varepsilon,\text{opt}}^{\text{common}}} v_{i,j} \cdot \Psi_{i,j}$$

Using these definitions we can write

$$\begin{aligned}
 \|f - f_{\varepsilon,d}\|_q &= \|f - f_{\varepsilon,\text{opt}} + f_{\varepsilon,\text{opt}} - f_{\varepsilon,d}\|_q \\
 &= \|f - f_{\varepsilon,\text{opt}} + (f_{\varepsilon,\text{opt}}^{\text{unique}} + f^{\text{common}}) - (f_{\varepsilon,d}^{\text{unique}} + f^{\text{common}})\|_q \\
 &\leq \|f - (f_{\varepsilon,\text{opt}} + f_{\varepsilon,d}^{\text{unique}})\|_q + \|f_{\varepsilon,\text{opt}}^{\text{unique}}\|_q
 \end{aligned} \tag{10}$$

Assuming that the adaptivity of the  $h$ -GSG method works perfectly, that is

$$\gamma_{i,j} = \|v_{i,j} \cdot \Psi_{i,j}\|_q \leq \varepsilon, \quad \forall (i,j) \notin P_{\varepsilon,d}^{\text{unique}}$$

then

$$\begin{aligned}
 \|f_{\varepsilon,\text{opt}}^{\text{unique}}\|_q &= \left\| \sum_{(i,j) \in P_{\varepsilon,\text{opt}}^{\text{unique}}} v_{i,j} \cdot \Psi_{i,j} \right\|_q \\
 &\leq \sum_{(i,j) \in P_{\varepsilon,\text{opt}}^{\text{unique}}} \|v_{i,j} \cdot \Psi_{i,j}\|_q \\
 &\leq \#(P_{\varepsilon,\text{opt}}^{\text{unique}}) \cdot \varepsilon
 \end{aligned} \tag{11}$$

By definition of the optimal interpolant

$$\|f - (f_{\varepsilon,\text{opt}} + f_{\varepsilon,d}^{\text{unique}})\|_q \leq \varepsilon \tag{12}$$

Setting  $N(\varepsilon) = \#(P_{\varepsilon,\text{opt}}^{\text{unique}})$  we arrive at the assertion.  $\square$

Theorem 1 states that the accuracy of the  $h$ -GSG interpolant is dependent on the number of points with cumulative  $\gamma_{i,j} < \varepsilon$  that are not in the approximation but have  $\gamma_{i,j} \approx \varepsilon$ . The exact number  $N(\varepsilon)$  of these points is dependent on the smoothness of the function being approximated. The smoother the function, that is the faster the hierarchical coefficients decay, the smaller  $N(\varepsilon)$  will be.

## 4 Numerical Study

In this section we investigate the performance of the proposed  $h$ -GSG method when applied to a number of numerical examples. We analyze convergence, with respect to the order of the local polynomial basis and the dimensionality for functions of varying smoothness. First we consider a set of two-dimensional functions which visually illustrates the effect of the choice of basis degree and the performance of local adaptivity. We then discuss the performance of different basis functions when applied to functions of differing regularity. The effect of the termination condition presented in Section 2.4 is also presented. Finally the utility of the proposed method is shown for high-dimensional approximation with hundreds of variables.

In the following we will consider the following four functions:

$$f_1(\xi) = \frac{1}{|0.3 - \xi_1^2 - \xi_2^2| + 0.1}, \quad \xi \in [0, 1]^2 \quad (13)$$

$$f_2^d(\xi) = \exp \left( - \sum_{i=1}^d c_i^2 (\xi_i - w_i)^2 \right), \quad \xi \in [0, 1]^d \quad (14)$$

$$f_3^d(\xi) = \exp \left( - \sum_{i=1}^d c_i |\xi_i - w_i| \right), \quad \xi \in [0, 1]^d \quad (15)$$

$$f_4^d(\xi) = \begin{cases} 0 & \text{if } \xi_1 > w_1 \text{ or } \xi_2 > w_2 \\ \exp(\sum_{i=1}^d c_i \xi_i) & \text{otherwise} \end{cases}, \quad \xi \in [0, 1]^d \quad (16)$$

Unless otherwise stated, the coefficients  $w_i = 0.5$ ,  $i = 1, \dots, d$ . The choice of  $c_i$  determines the effective dimensionality of the function and is defined differently for each problem. Although smooth, the mixed derivatives of the Gaussian function  $f_2^d$  can become large and thus degrade performance if not compensated for by appropriate adaptivity. The discontinuities in functions  $f_3^d$  and  $f_4^d$  also degrade, with increasing magnitude, the efficiency of isotropic methods and subsequently can highlight the strengths and weaknesses of any interpolation method.

In the following we will analyze convergence with respect to the following measures:

$$\begin{aligned} \varepsilon_{\ell^\infty} &= \max_{i=1, \dots, N} |f(\xi_i) - g(\xi_i)| \\ \varepsilon_{\ell^2} &= \left( \frac{1}{N} \sum_{i=1}^N |f(\xi_i) - g(\xi_i)|^2 \right)^{1/2} \end{aligned}$$

where  $f$  and  $g$  are the true function and approximation respectively. In all the following examples  $N = 1000$ . Error in the quadrature rule  $I_{\text{approx}}$  is also considered and measured by

$$\varepsilon_{\text{integral}} = \frac{I_{\text{approx}} - I_{\text{exact}}}{I_{\text{exact}}}$$

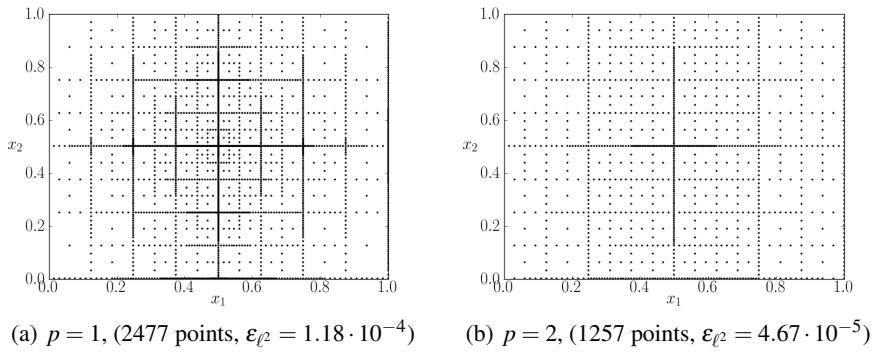
where  $I_{\text{exact}}$  is the exact integral. Unless otherwise stated this value is calculated analytically.

#### 4.1 A two-dimensional example

Let us first consider two low dimensional functions, (13) and (15), defined on the unit hypercube  $[0, 1]^2$ . Figure 3 depicts the grids generated when the proposed method is applied to the piecewise continuous function (15) using linear basis functions (a) and quadratic basis functions (b). This function has discontinuities in its first derivatives along  $\xi_1 = 0.5$  and  $\xi_2 = 0.5$ . The linear  $h$ -GSG grid concentrates grid points around the rapidly varying region associated with the discontinuous change in the derivative information. In comparison the quadratic basis requires significantly less function evaluations. The quadratic basis is able to obtain second order convergence in each

of the four smooth quadrants whilst still approximating well at the discontinuities. Here  $c_i = 10/2^{i+2}$ , and the absolute error criterion (8) is used with  $\varepsilon = 10^{-6}$ .

The function (15) possesses discontinuities that lie along the axial directions. Let us now consider function (13) which possesses a singularity that passes through both axial directions. Figure 4 depicts the grids obtained using a tolerance of  $\varepsilon = 10^{-6}$  and linear (a) and quadratic basis functions (b). The linear  $h$ -GSG grid concentrates grid points around the rapidly varying region associated with the discontinuous change in the derivative information. Unlike the previous example it is now unclear whether the use of the quadratic basis function results in increased efficiency. The accuracy of the linear basis functions is higher ( $\varepsilon_{\ell^2} = 3.19 \cdot 10^{-3}$ ) than when the quadratic basis is used ( $\varepsilon_{\ell^2} = 1.15 \cdot 10^{-2}$ ) but the linear method requires almost three times as many points. The effect of varying the degree of the local basis is discussed in the following section.

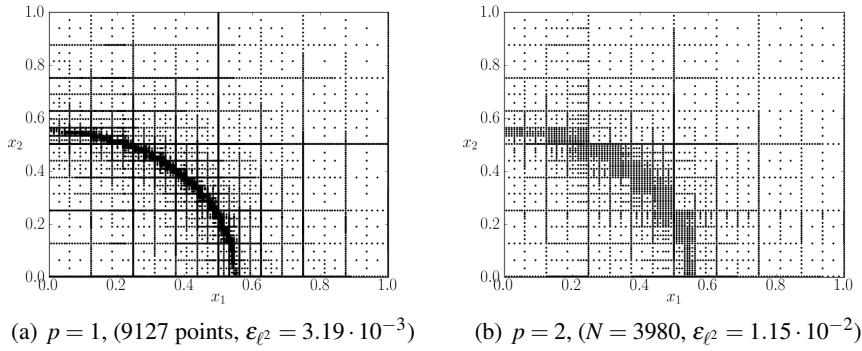


**Fig. 3.** The adaptive grids obtained using basis of varying degree. Here the error criterion (8) is used with  $\varepsilon = 10^{-6}$ .

#### 4.2 Increasing the Degree of the Local Polynomial Basis

Let us now consider some moderate-dimensional integrals ( $d = 10$ ) and discuss the effect of the degree of the local polynomial basis on the efficiency of the proposed method. Setting  $c_i = 1/2^{i+2}$ , Figure 5 compares the rates of convergence with respect to the tolerance  $\varepsilon$  when the proposed method is applied to the three test functions (14)-(16). Figure 5 (a) illustrates convergence with respect to the  $\varepsilon_{\ell^\infty}$  measure for the smooth function  $f_2^{d=10}$ . In this case the effect of the higher-degree basis is clearly evident. The quadratic basis provides drastic improvement over the standard piecewise-linear basis and the quartic basis provides a further increase in efficiency.

This result is mirrored when  $h$ -GSG is applied to the piecewise-continuous function  $f_3^{d=10}$  (15) (Figure 5 (b)). However when a jump discontinuity is present (function (16)  $f_4^{d=10}$ ) the performance of the quartic basis functions is reduced (Figure 5 (c)).



**Fig. 4.** The adaptive grids obtained using basis of varying degree. Here the error criterion (8) is used with  $\varepsilon = 10^{-6}$ .

The quadratic basis significantly increases the accuracy of the *h*-GSG method for smooth and discontinuous functions. Higher order basis functions  $p > 2$  provide further increases in the rate of convergence obtained for smooth problems, but performance is degraded for discontinuous problems. These results are reproduced when higher-dimensional realizations of these functions are considered.

#### 4.3 Efficient Termination of *h*-GSG

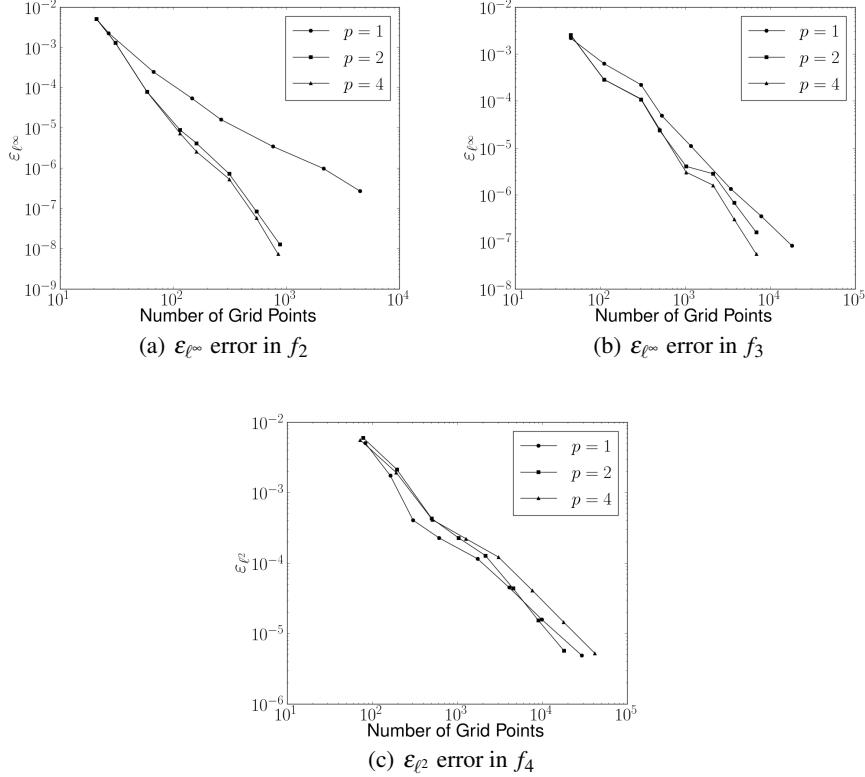
In Section 2.4 we proposed that the efficiency of the generalised sparse grid algorithm can be improved by only adding admissible indices with  $r_i > \varepsilon$  to the active index set. Here we substantiate that claim. The *h*-GSG method discussed here and throughout this manuscript implements this modification.

Again consider the three test functions (14)-(16). But now let us investigate performance when  $d = 100$  and

$$c_i = \lambda \exp\left(-\frac{35 \cdot i}{d}\right), \quad i = 1, \dots, d$$

where the parameter  $\lambda$  which controls the effective dimensionality of the function. Figures 6 illustrates the difference between the proposed method with and without the modification proposed in Section 2.4. Specifically the figures depict the error in the sparse grid interpolant as the algorithm evolves. The modification results in substantial improvement when applied to the smooth (not depicted) and piecewise-continuous (Figure 6 (a)) functions. The unmodified algorithm adds many points corresponding to grid indices with  $r_i < \varepsilon$  and which contribute to the interpolation error. The modification limits the number of unimportant points.

When applied to the discontinuous function (Figure 6 (b)) the modification results in lower accuracy than when the unmodified algorithm is used. The unmodified algorithm continues to add points belonging to grids with  $r_i < \varepsilon$  and which contribute little to the integral of the function yet still significantly influence the accuracy of the

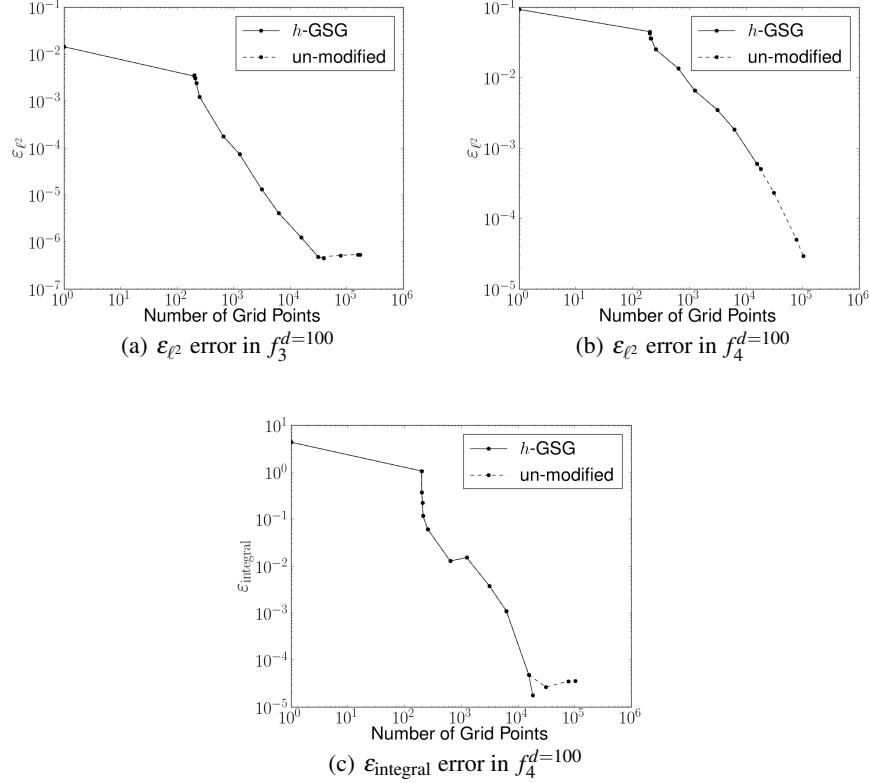


**Fig. 5.** Error in the interpolants of  $f_2$ ,  $f_3$ , and  $f_4$  for  $p \in \{1, 2, 4\}$ ,  $\varepsilon = 10^{-8}$  and  $d = 10$ .

interpolant. The points belonging to  $r_i < \varepsilon$  mainly reside around the discontinuity. As the level of refinement increases, the contribution of these points to the integral decreases yet their effect on the interpolant may not. This effect is illustrated in Figure 6 (c) which depicts the decrease in the error of the integral approximation with and without the modification when  $h$ -GSG is applied to the discontinuous function. Here it is clear that adding grids with  $r_i < \varepsilon$  has little effect on the accuracy of the integral approximation. Also note that the effect of the termination condition decreases when the dimensionality  $d$  is small.

#### 4.4 High-Dimensional Interpolation

In Section 4.3 we used  $h$ -GSG for interpolation for a set of 100-dimensional problems. In this section we show that the proposed method can be applied to much higher-dimensional problems. Again consider the discontinuous function (16) which is the most difficult function to approximate of the three test functions used thus far. Again let



**Fig. 6.** The evolution of the error in the interpolant when applied to the test functions with and without the proposed termination condition. Here  $d = 100$  and  $\varepsilon = 10^{-6}$ . Initially the results are visually identical. Differences only occur towards the end of the algorithm.

$$c_i = \lambda \exp\left(-\frac{35 \cdot i}{d}\right) \quad (17)$$

Table 1 shows the number of function evaluations required to approximate  $f_4$  and the resulting relative error in the approximated integral when  $\varepsilon = 10^{-4}$ , quadratic basis functions are used, and  $\lambda = 1$ . An analytical expression for the integrand can be obtained easily due to the exponential nature of the function. Arbitrary precision arithmetic was used to evaluate the numerical value of the reference integrals. Due to the large range of values that  $f_4$  can take in high dimensions we use the relative error indicators

$$r_i = \left| \frac{\sum_{j \in B_i} v_{i,j} \cdot w_{i,j}}{w_{0,0} \cdot v_{0,0}} \right|, \quad \gamma_i = \left| \frac{v_{i,j} \cdot w_{i,j}}{w_{0,0} \cdot v_{0,0}} \right|$$

to respectively guide difference space selection and local refinement. The *h*-GSG algorithm is terminated when

$$\left| \frac{\sum_{\mathbf{i} \in \mathcal{J}} r_{\mathbf{i}}}{w_{0,0} \cdot v_{0,0}} \right| < \varepsilon$$

where  $r$  is the global error indicator used in Algorithm 1. An error of the order  $10^{-2}$  is achieved for up to 700 dimensions using less than 300,000 function evaluations.

The accuracy of the integral approximation decays with increasing dimensionality. This is likely caused by the particular error indicators ( $\gamma_{i,j}$  and  $r_i$ ) used to guide adaptivity. At the moment a point  $\xi_{i,j}$  is refined if  $\gamma_{i,j} \geq \varepsilon$  and a grid index  $\mathbf{i}$  is flagged for refinement only if  $r_i \geq \varepsilon$ . This approach works well when  $d < 400$  but could be improved upon when the dimensionality is higher. By excluding points if they have  $\gamma_{i,j}$  less than the desired accuracy  $\varepsilon$  we are potentially ignoring a significant number of points whose combined contribution to the integral is greater than  $\varepsilon$ . As the dimensionality increases more and more points will be excluded from consideration thereby causing the accuracy of the approximant to decrease. This remark is consistent with Theorem 1 which states that the accuracy of the  $h$ -GSG approximation is dependent on the number of points with  $\gamma_{i,j}$  close to  $\varepsilon$ . As the number of these points increases the accuracy of the approximation decreases.

The decrease in accuracy depicted in Table 1 could be addressed by utilising more appropriate error criteria than those used here. The construction of efficient and robust error indicators is problem dependent and must be based upon the properties of the function under consideration. If no information on the function is available, we have shown that the error indicators used here will still perform well.

**Table 1.** Errors in the  $h$ -GSG approximation of (16) for  $d = 100$  to 700.  $\varepsilon = 10^{-5}$

$d$	$N$	$\varepsilon_{\text{integral}}$
100	3,376	$3.81 \cdot 10^{-4}$
200	12,488	$1.67 \cdot 10^{-3}$
300	31,533	$1.71 \cdot 10^{-4}$
400	62,404	$8.44 \cdot 10^{-5}$
500	109,356	$4.57 \cdot 10^{-3}$
600	176,842	$7.97 \cdot 10^{-3}$
700	269,665	$1.68 \cdot 10^{-2}$

Here we note that the rate of convergence of the  $h$ -GSG method is governed by the implicit weighting of the importance of each dimension. In this case the importance is controlled by the coefficients  $c_i$ . To illustrate this dependence, Table 2 shows the efficiency of  $h$ -GSG as the magnitude of the coefficients is increased. As the “importance” of each dimension increases the number of sub-dimensional components increases. In this case an increase in  $c_i$  also increases the function variability which also requires additional points to achieve a set level of accuracy.

**Table 2.** Errors in the  $h$ -GSG approximation of (16) for increasing dimension importance  $\varepsilon = 10^{-6}$ . Importance is increased by increasing  $\lambda$  in Equation (17).

$\lambda$	$N$	$\varepsilon_{\text{integral}}$
1	9,226	$1.66 \cdot 10^{-4}$
2.5	34,977	$2.96 \cdot 10^{-5}$
5	175,201	$6.53 \cdot 10^{-4}$
7.5	659,368	$1.93 \cdot 10^{-3}$

## 5 Conclusion

This paper presented an  $h$ -adaptive generalised sparse grid ( $h$ -GSG) method for interpolating high-dimensional functions with discontinuities. The proposed algorithm extends and improves upon existing approaches by combining the strengths of the generalised sparse grid algorithm and hierarchical surplus-guided  $h$ -adaptivity.

The underlying generalised sparse grid algorithm greedily selects the subspaces that contribute most to the variability of a function. The hierarchical surplus of the points within each subspace is used as an error criterion for  $h$ -refinement with the aim of concentrating computational effort within rapidly varying or discontinuous regions. This approach limits the number of points that are invested in ‘unimportant’ subspaces and regions within the high-dimensional domain.

A high-degree basis is used to obtain a high-order method that, given sufficient smoothness, performs significantly better than the traditional piecewise-linear basis. When discontinuities are present in the function surface or its derivatives, performance deteriorates. However, it was shown numerically that even in such situations the quadratic basis will still result in higher-rates of convergence than that achieved by using piecewise-linear interpolation.

Often the importance of function variables are governed by natural yet unknown weights. In these cases, the proposed method can utilise this implicit weighting to determine and restrict effort to the effective dimension of the model. This property allows the  $h$ -GSG method to be applied to non-smooth functions with hundreds of variables.

## References

1. Barthelmann, V., Novak, E., Ritter, K.: High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics* **12**, 273–288 (2000)
2. Bungartz, H., Dirnstorfer, S.: Multivariate quadrature on adaptive sparse grids. *Computing* **71**(1), 89–114 (2003). DOI [10.1007/s00607-003-0016-4](https://doi.org/10.1007/s00607-003-0016-4)
3. Bungartz, H.J.: Finite elements of higher order on sparse grids. Ph.D. thesis, Institut für Informatik, TU München (1998)
4. Bungartz, H.J., Griebel, M.: Sparse grids. *Acta Numerica* **13**, 147–269 (2004)
5. Gerstner, T., Griebel, M.: Numerical integration using sparse grids. *Numerical Algorithms* **18**(3-4), 209–232 (1998)

6. Gerstner, T., Griebel, M.: Dimension-adaptive tensor-product quadrature. *Computing* **71**(1), 65–87 (2003). DOI [10.1007/s00607-003-0015-5](https://doi.org/10.1007/s00607-003-0015-5)
7. Griebel, M.: Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing* **61**(2), 151–179 (1998). DOI <http://dx.doi.org/10.1007/BF02684411>
8. Griebel, M., Holtz, M.: Dimension-wise integration of high-dimensional functions with applications to finance. *Journal of Complexity* **26**(5), 455–489 (2010). DOI DOI: [10.1016/j.jco.2010.06.001](https://doi.org/10.1016/j.jco.2010.06.001). URL <http://www.sciencedirect.com/science/article/B6WHX-50F8BT9-1/2/b1c278f2e0e3ad9714b17aee72497c74>. SI: HDA 2009
9. Ma, X., Zabaras, N.: An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics* **228**, 3084–3113 (2009)
10. Ma, X., Zabaras, N.: An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. *Journal of Computational Physics* **229**(10), 3884–3915 (2010). DOI [10.1016/j.jcp.2010.01.033](https://doi.org/10.1016/j.jcp.2010.01.033). URL <http://dx.doi.org/10.1016/j.jcp.2010.01.033>
11. Pflüger, D.: Spatially adaptive sparse grids for high-dimensional problems. Dissertation, Institut für Informatik, Technische Universität München, München (2010). URL <http://www5.in.tum.de/pub/pflueger10spatially.pdf>
12. Pflüger, D., Peherstorfer, B., Bungartz, H.J.: Spatially adaptive sparse grids for high-dimensional data-driven problems. *Journal of Complexity* **26**(5), 508–522 (2010). DOI [10.1016/j.jco.2010.04.001](https://doi.org/10.1016/j.jco.2010.04.001). URL <http://www.sciencedirect.com/science/article/pii/S0885064X10000257>
13. Smolyak, S.: Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Math. Dokl.* **4**, 240–243 (1963)